一种基于速率的组播拥塞控制机制*

任立勇** 卢显良

（电子科技大学计算机科学与工程学院 成都 610054）

【摘要】 分析了发送者驱动和接收者驱动的组播拥塞控制的优缺点，提出了一种基于速率，由接收者和发送者混合驱动的层次型组播拥塞控制机制 RBMCC。在 RBMCC 中接收者计算本地丢失率，中间节点聚集所有来自其直接子节点的反馈报文，发送方通过最终的反馈报文计算 TCP 友好发送速率，实现对网络拥塞的快速、准确响应。仿真试验证明，RBMCC 具有良好的可伸缩性与满意的公平性。

关 键 词 组播； 拥塞控制； 基于速率； TCP 友好

中图分类号 TP302

端到端拥塞控制机制对 Internet 的鲁棒性和稳定性至关重要。目前 Internet 上较多使用基于窗口的拥塞控制算法对网络拥塞进行响应和避免网络崩溃。故 Internet 的成功很大程度上得益于 TCP 拥塞控制算法的不断改进[1~3]。

组播是一种为优化使用网络资源而产生的技术。它通常使用在点对多工作方式下的应用中，如软件分发、视频会议、交互式仿真等。组播的引入虽然节约了网络资源，减少了网络管理费用，但同时也给 Internet 带来了许多潜在的问题，如组播安全、拥塞控制、组播路由、错误恢复问题等，而组播的拥塞控制是最重要、最难解决、最令人关注的问题之一。如果组播应用不能对网络拥塞作出正确响应，将会给 Internet 带来比单点投递应用产生的拥塞更为严重的影响。其主要原因是：1）组播投递流可能沿着多点投递树广泛分布于整个 Internet；2）组播流的接收者具有异构性，每个接受者的处理能力不同，其分组投递路径也可能有不同的带宽和差错特性；3）组播发送方需处理比单点投递多得多的反馈报文，对这些报文如果不加以处理，不仅可能淹没发送方，而且其本身也是对网络资源的一种极大的开销。

本文提出了一种基于速率的组播拥塞控制机制(RBMCC)。RBMCC 利用已有的 IP 组播技术将组播接收者构建成树型逻辑结构，不仅较好地解决了拥塞控制的可伸缩性问题，同时能有效地遏制反馈内陷。同时，通过采用稳定状态下的 TCP 吞吐量模型作为组播的速率调节依据，RBMCC 具有较好的 TCP 友好特性。

1 RBMCC 拥塞控制模型

本文提出的 RBMCC 拥塞控制模型如图 1 所示，该模型具有几个新的特点。首先，模型采用基于接收方与发送方的混合式驱动。在发送方驱动模式中，每个接收者向发送方发送 ACK 报文，发送方维护大量的接收者状态信息及定时器信息，承担拥塞控制的绝大部分任务[4]。这种方式好处是发送方能及时了解整个组播组的状况，并作出响应，但其最大的缺点是发送方开销太大，容易成为网络瓶颈，组播的可伸缩性较差。在接收方驱动模式中，每个接收者依据其接收报文的情况(丢失率、RTT 等)，计算其“最佳”的接收速率，当出现报文丢失、损坏时向发送方发送 NAK 报文。这种方式的好处是将大量的计算任务分配到各个接收者，减轻了发送方的负担，组播组的可伸缩性较好。同时由接收者来计算其接收速率，其计算准确度较高。但纯接收方驱动可能导致发送方死锁，因为发送方不能确定所有接收者何时正确接收数据报文[5]。为此，本文采用了发送方与接收方的混

2001 年 8 月 28 日收稿
* 信息产业部生产发展基金资助项目。
** 男 30 岁 博士生

合式驱动模式，即接收者每接收一个报文，均计算其丢失率，并向发送方发送特定的反馈报文，而发送方根据接收者反馈报文计算 RTT，计算并调整发送速率。

其次，为避免反馈爆炸，增强组播拥塞控制的可伸缩性，本文采用了层次型、多级的反馈报文聚集的网络模型。即中间节点对其直接子节点的反馈报文进行聚合，并向其上一级发送聚集反馈报文(AAK)。同时，中间节点还可以担负本地恢复错误报文的任务(本文不作讨论)。

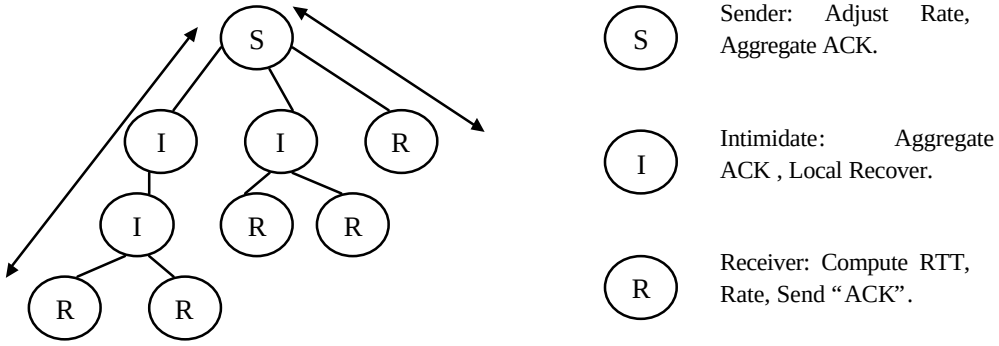

图 1 RBMCC 网络模型

为保证组播发送方对网络的拥塞作出正确的响应，并能与竞争流(TCP 流)公平共享瓶颈链路带宽，RBMCC 采用了稳定状态下的 TCP 流量模型作为其速率调节公式[6]，即

$$T = \frac{M}{\mathrm{RTT}\sqrt{2p/3} + T_{RTO}\left(3\sqrt{3p/8}\right)p\left(1+32p^2\right)} \tag{1}$$

式中 T 为发送速率；M 为报文长度；RTT 为数据报文的往返时延；T_{RTO} 为式(1)重传超时值；p 为报文丢失率。该流量模型的有效性验证在许多文献中都有描述[7]，需要指出的是，式(1)中往返时延 RTT 及丢失率 p 在组播协议中需作特殊的处理，以保证各接收者间能协同工作。

2 RBMCC 实现细节

2.1 接收方

RBMCC 拥塞控制中接收方承担着计算从发送方到自身所经历路径的报文丢失率、当前接收速率，并发送反馈报文。主要过程如下：

1）每当接收到一个报文，计算报文丢失率，设为 p；

2）构造反馈报文，主要包括以下几部分：确认报文号 AckSeq，时间戳 TimeStamp，接收者编 ReceId，丢失率 p，其中 TimeStamp 为当前接收报文的发送时间(由发送方标记)；

3）向其父节点发送 2）构造的反馈报文。

分组丢失率作为评判网络拥塞状况的一个重要参数，其估计值的准确性对传输速率的调节相当关键。本文采用随网络状态变化的动态时间尺度计算分组丢失率。即设丢失间隔 s 为两个相邻丢失事件间正确接收的报文数，则 $p=1/s$。为了保证丢失率的估计值更加平滑，同时在一定程度上反应历史丢包事件，对丢失率进行了具有低通滤波效果的指数加权移动平均，即

$$\tilde{s} \leftarrow (1-h)\ \tilde{s} + h\ \ s_{rec} \qquad h \in [0,1] \tag{2}$$

式中 $\tilde{s}$ 为平均丢失间隔；s_{rec} 为最近丢失间隔。为更及时反应网络拥塞状况，h 的取值可尽量大些(仿真试验验证，当 h 取[0.6~0.8]，估计效果较好)。

2.2 中间节点

RBMCC 拥塞控制中中间节点承担着对其直接子节点的反馈报文进行聚合，并向上一级中间节点(或发送方)发送聚集报文 AAK 的任务，同时，对于可靠组播，中间节点还承担着当其子节点有报文丢失时，完成本地恢复的任务。即，中间节点主要完成以下几个任务：

1) 中间节点为每个下游接口保存几个本地变量，并分别设置初始值：I_AckSeq=0，I_ReceId=0，

I_TimeStamp，I_p=0。

2) 当接收到一个从下游传来的 AAK 时，利用 AAK 中值更新相应接口的变量。

3) 构造一个新的 AAK_n，构造规则如下：

(1) 从所有接口的本地变量中选择最小的 I_AckSeq 赋给 $AckSeq_n$，并取相应的 I_TimeStamp，I_ReceId 分别赋予 $TimeStamp_n$、$ReceId_n$；

(2) 从所有接口的本地变量中选择最大的 I_p 赋予 p_n。

4) 当 $AckSeq_n \neq 0$ 时，满足以下条件之一时，中间节点向上一级发送新构造的 AAK_n。

(1) $ReceId_n \neq$ I_ReceId；

(2) $ReceId_n$=I_ReceId，并且 $AckSeq_n$>I_AckSeq。

需要指出的是，通过上述的聚集运算可以保证每个数据报文在组播树的每个子树上最多发送一次反馈报文，并且都是由最后发送 ACK 的接收者来触发，这样就可以避免组播的归零问题，同时也极大地避免了反馈爆炸问题，同时聚集报文中还包括其下游链路的最小速率与最大丢失率。

2.3　发送方

RBMCC 中发送方主要包含聚集模块与速率调整模块两个模块。其中聚集模块采用与中间节点完全一致的聚集算法对来自其直接子节点的反馈报文(AAK 或 ACK)进行聚合，并将聚集报文送交速率调整模块。

速率调整模块的主要任务是根据反馈聚集报文计算理想发送速率，并调整发送速率，即，速率调整模块每收到一个 AAK，按以下步骤调整速率：

1) 计算 RTT, 令 rtt=time-TimeStamp，其中 time 为收到 AAK 时当前时间，TimeStamp 为 AAK 中时间戳（即该反馈报文对应的数据报文的发送时间）。为过滤因为偶然因素引起的 RTT 剧变，平滑估计往返时延，本文采用指数加权移动平均 EWMA 计算往返时延 RTT，即

$$\mathrm{RTT} \leftarrow (1-m)\mathrm{RTT} + m \ \ \mathrm{rtt} \qquad m \in [0,1] \tag{3}$$

2) 令超时值 T_{RTO}=4RTT，这种估算与 TCP Reno 一致。

3) 根据 2)计算当前理想的发送速率 T。

4) $|T - T_{act}| > \Delta$时，令实际发送速率 $T_{act} \leftarrow T$，否则，发送速率不变。其中 Δ为可配置参数，作这样的处理的目的，主要是考虑不同组播应用对速率振荡要求不同。如实时媒体分发就要求尽量避免剧烈速率振荡，此时可将 Δ设置略微大一些，而对于可靠组播，则可将 Δ设置小一些，甚至为 0。

尽管本文没有考虑报文超时重发的问题，但由于超时说明在组播树中已发送严重的拥塞。因此，RBMCC 中发送方维护着一个超时变量 T_{RTO}，当超时事件发生，将发送速率调整到组播的初始速率(对可靠组播其初始速率为一个报文长度，而实时媒体组播其初始速率为最低服务质量要求的发送速率)。

3　仿真试验

为验证 RBMCC 的有效性，即具有良好的可伸缩性、TCP 友好等＝1000 ，已在 NS-2 中实现了 RBMCC。同时也进行了各种网络情况下的仿真试验，限于篇幅，本文仅给出了一个试验结果：当 RBMCC 与 TCP 流竞争瓶颈带宽时，RBMCC 的公平性验证。为简单起见，本文选用实时媒体组播作为 RBMCC 的应用对象，即没有考虑超时重发的问题。为此，在试验中取 η=0.65，μ=0.3，Δ=1 000 byte。

图 2 为验证 RBMCC 公平性(即 TCP 友好)的网络拓扑结构。其中 S_1 为 RBMCC 流发送方，R_1，R_2 为 RBMCC 流两个接收者，S_2 为 TCP 流发送方，R_3 为 TCP 流接收者。RBMCC 流初始发送速率为 10 kbps，试验持续 10 s。其中 TCP 流为一大文件传输流，在第 2 s 时开始，第 4 s 中断，第二次启动时间为第 7s，图 3 为该试验结果图。当 RBMCC 流与 TCP 流竞争瓶颈带宽时(2～4 s，7～10 s)RBMCC 流略微比 TCP 流多占用 30%左右的带宽，这主要是因为 RBMCC 流没有考虑超时重传所致，但 RBMCC 流基本上能与 TCP 流公平共享瓶颈带宽。另外，在时刻 6 s 与 8 s 附近，RBMCC

流速率回调到初始发送速率，这是因为出现了反馈超时。

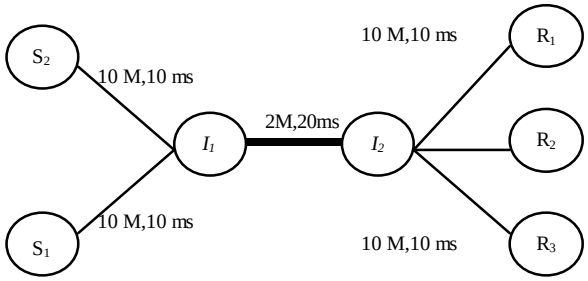

图 2 TCP-Friendly 试验拓扑

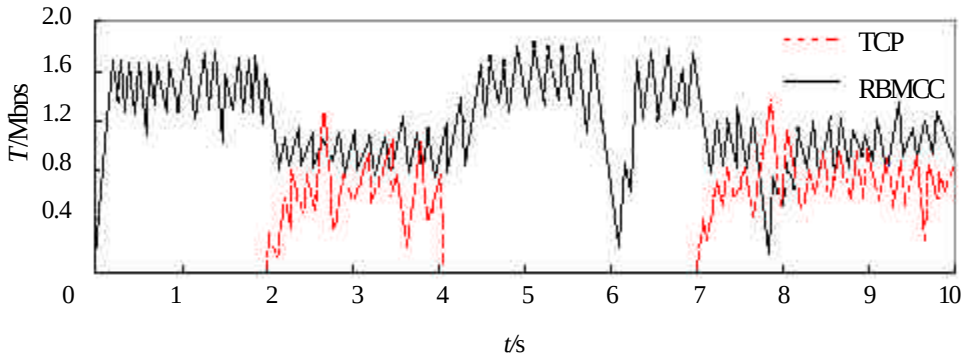

图 3 RBMCC 的 TCP－Friendly 试验

4 结 束 语

组播拥塞控制是设计组播协议的最重要问题之一，伸缩、公平的组播拥塞机制也是组播协议在互联网络中得以应用的关键前提。本文提出了一种基于速率、接收者与发送方混合驱动的组播拥塞控制机制 RBMCC，由接收者计算本地丢失率、中间节点聚集反馈报文、发送方计算 TCP 友好发送速率，并最终实现对网络拥塞的快速、准确响应。仿真试验证明，RBMCC 具有良好的可伸缩性与满意的公平性。

需要特别指出的是，由于 RBMCC 没有考虑超时重传问题，因此与 TCP 流共享带宽时，还不能达到“绝对”公平，表现出一定的侵略性。另外，现有的 RBMCC 还没有在互联网络中得以验证，有待作进一步研究。

参 考 文 献

1 Allman M, Paxson V, Stevens W. TCP congestion control. request for comments: 2581, April 1999.

2 Wang S, Li L M. Asymptotic analysis of a linear closed-loop congestion control scheme. Journal of University of Electronic Science and Technology of China, 2000.29(4): 450~456[王 晟.李乐民.一种线性闭环拥塞控制方案的渐近性能分析.电子科技大学学报, 2000.29(4): 450~456]

3 Tu X D, Li L M. Research on simulation of packet fair oueueing algorithms. Journal of University of Electronic Science and Technology of China, 2000.29(4): 440~444.[涂晓东.李乐民.分组公平排队算法的仿真研究.电子科技大学学报.2000.29(4): 440~444]

4 Pingali S, Towsley D, Kurose,J F. A comparison of sender-initiated and receiver-initiated reliable multicsast protocols. In Proceedings of the Sigmetrics Conference on Measurement and Modeling of Computer Systems, 1994: 221-230.

5　Maihofer C. A bandwidth analysis of reliable multicast transport protocols. In Proceedings of the Second International Workshop on Networked Group Communication(NGC 2000), 2000: 15-26.

6　Padhye J, Firoiu V, Towsley, *et al.* Modeling TCP throughput: a simple model and its empirical validation. Umass-CMPSCI Technical Report TR 98-008, Feb 1998.

7　任立勇,卢显良.一种基于方程的多媒体实时流拥塞控制机制.计算机科学,2001. 23(6):60~63.

A Rate-Based Congestion Control Mechanism for Multicasts

Ren Liyong　　Lu Xianliang

(College of Computer Science and Engineering, UEST of China　Chengdu　610054)

Abstract Congestion control is key problem in designing multicast protocols. The merits and shortcomings of sender-drivered and receiver-drivered in congestion control are analyzed respectively. Then a congestion control mechanism for multicast, RBMCC that is rate-based and drivered mixedly by sender and receiver, is presented. In RBMCC, each receiver estimate its packet loss rate respectively and send ACKs to its parent, intermediate nodes aggregated these ACKs that come from all of children. TCP-friendly rate computed by sender provide timely and accurate response to the network congestion. The results of experimental simulation show that RBMCC is quite flexible with satisfactory equality.

Key word　multicast;　congestion control;　rate-based;　TCP-friendly

---

·科研成果介绍·

8 mm 微波前端模块化技术

主研人员：徐锐敏　薛良金　延 波等

8 mm 微波前端模块化技术研制的样机，采用开关型接收原理，将毫米波信号通过开关切换进行了频域分割，下变频到微波频段，以现有的频段接收机作为终端。解决了毫米波放大器的宽带低噪声、混频器的低交调输出和系统的电磁兼容等关键技术。

专家系统网络重组技术

主研人员：郭 伟　尹道素　田永春等

专家系统网络重组技术提出了网络可靠性优化设计模型、网络重组优化算法和网络初始化及稳定运行后的动态优化设计算法。建立了军用移动通信网可靠性、抗毁性、总代价评估模型；开展了网络初始规划设计和路由控制设计的研究，提出了渐进寻优的递推启发式算法。

·甬 江·